

# Asymptotic normality through factorial cumulants and partitions identities\*

Konstancja Bobecka<sup>†</sup>, Paweł Hitczenko<sup>‡</sup>, Fernando López-Blázquez<sup>§</sup>, Grzegorz Rempala<sup>¶</sup>, Jacek Wesolowski<sup>||</sup>

June 24, 2011

## Abstract

In the paper we develop an approach to asymptotic normality through factorial cumulants. Factorial cumulants arise in the same manner from factorial moments, as (ordinary) cumulants from (ordinary) moments. Another tool we exploit is a new identity for "moments" of partitions of numbers. The general limiting result is then used to (re)derive asymptotic normality for several models including classical discrete distributions, occupancy problems in some generalized allocation schemes and two models related to negative multinomial distribution.

## 1 Introduction

Convergence to normal law is one of the most important phenomena of probability. As a consequence, a number of general methods, often based on transforms of the underlying sequence have been developed as techniques for establishing such convergence. One of these methods, called the method of moments, rests on the fact that the standard normal random variable is uniquely determined by its moments and that for such random variable  $X$  if  $(X_n)$  is a sequence of random variables having all moments and  $\mathbb{E} X_n^k \rightarrow \mathbb{E} X^k$  for

---

\*KB, FLB and JW were partially supported by Spanish Ministry of Education and Science Grant MTM2007-65909. Part of work of KB and JW was carried while they were visiting University of Seville in March/April 2009. PH was partially supported by NSA Grant #H98230-09-1-0062. Part of work of PH was carried out while he was at the Institute of Mathematics of the Polish Academy of Sciences and the Technical University of Warsaw in the Fall of 2010. GR and JW were partially supported by NSF Grant DMS0840695 and NIH Grant R01CA152158. Part of work of GR was carried when he was visiting Warsaw University of Technology in July 2009. Part of work of JW was carried out while he was visiting Medical College of Georgia in January/February 2009.

<sup>†</sup>Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska, Warszawa, Poland e-mail: bobecka@mini.pw.edu.pl

<sup>‡</sup>Department of Mathematics, Drexel University, Philadelphia, USA, e-mail: phitczenko@math.drexel.edu

<sup>§</sup>Facultad de Matemáticas Universidad de Sevilla, Sevilla, Spain, e-mail: lopez@us.es

<sup>¶</sup>Department of Biostatistics, Medical College of Georgia, Augusta, USA, e-mail: grempala@mccg.edu

<sup>||</sup>Wydział Matematyki i Nauk Informacyjnych, Politechnika Warszawska, Warszawa, Poland, e-mail: wesolo@mini.pw.edu.pl

all  $k = 1, 2, \dots$ , then  $X_n \xrightarrow{d} X$  - see, e.g. [22, Theorem 2.22] or [2, Theorem 30.2]. Here and throughout the paper we use “ $\xrightarrow{d}$ ” to denote the convergence in distribution.

Since moments are not always convenient to work with, one can often use some other characteristics of random variables to establish the convergence to normal law. For example in one classical situation we consider a sequence of cumulants (we recall the definition in the next section) rather than moments. On the one hand, since the  $k$ th cumulant is a continuous function of the first  $k$  moments, to prove that  $X_n \xrightarrow{d} X$  instead of convergence of moments one can use convergence of cumulants. On the other hand, all cumulants of the standard normal distribution are zero except of the cumulant of order 2 which equals 1. This often makes establishing the convergence of cumulants of  $(X_n)$  to the cumulants of the standard normal random variable much easier. We refer the reader to e.g. [8, Section 6.1] for more detailed discussion.

In this paper we develop an approach to asymptotic normality that is based on factorial cumulants. They will be discussed in the forthcoming section, here we just indicate that factorial cumulants arise in the same manner from factorial moments, as (ordinary) cumulants from (ordinary) moments. The motivation for our work is the fact that quite often one encounters situations in which properties of random variables are naturally expressed through factorial (rather than ordinary) moments. As we will see below, this is the case, for instance, when random variables under consideration are sums of indicator variables.

In developing our approach we first provide a simple and yet quite general sufficient condition for the central limit theorem in terms of factorial cumulants (see Proposition 1 in the next section). Further, as we will see in Theorem 3 below, we show that the validity of this condition can be verified by controlling the asymptotic behavior of factorial moments. This limiting result will be then used in Section 5 to (re)derive asymptotic normality for several models including classical discrete distributions, occupancy problems in some generalized allocation schemes (gas), and two models related to a negative multinomial distribution; they are examples of what may be called generalized inverse allocation schemes (gias). Gas was introduced in [13] and we refer the reader to chapters in books [14, 15] by the same author for more details, properties, and further references. The term “generalized inverse allocation schemes” does not seem to be commonly used in the literature; in our terminology the word “inverse” refers to inverse sampling, a sampling method proposed in [6]. A number of distributions with “inverse” in their names reflecting the inverse sampling are discussed in the first ([10]) and the fourth ([9]) volume on probability distributions

We believe that our approach may turn out to be useful in other situations when the factorial moments are natural and convenient quantities to work with. We wish to mention, however, that although several of our summation schemes are, in fact, gas or gias we do not have a general statement that would give a reasonably general sufficient conditions under which a gas or a gias exhibits the asymptotic normality. It is perhaps, and interesting question, worth further investigation.

Aside from utilizing factorial cumulants, another technical tool we exploit is an identity for “moments” of partitions of natural numbers (see Proposition 2 in Section 3). As far as we know this identity is new and may be of independent interest to combinatorics community. As of now, however, we do not have any combinatorial interpretation, neither for its validity nor for its proof.

## 2 Factorial cumulants

Let  $X$  be a random variable with the Laplace transform

$$\phi_X(t) = \mathbb{E} e^{tX} = \sum_{k=0}^{\infty} \mu_k \frac{t^k}{k!}$$

and the cumulant transform

$$\psi_X(t) = \log(\phi_X(t)) = \sum_{k=0}^{\infty} \kappa_k \frac{t^k}{k!} .$$

Then  $\mu_k = \mathbb{E} X^k$  and  $\kappa_k$  are, respectively, the  $k$ th moment and the  $k$ th cumulant of  $X$ ,  $k = 0, 1, \dots$ . The Laplace transform can be also expanded as

$$\phi_X(t) = \sum_{k=0}^{\infty} \nu_k \frac{(e^t - 1)^k}{k!} , \quad (2.1)$$

where  $\nu_k$  is the  $k$ th factorial moment of  $X$ , that is  $\nu_0 = 1$  and  $\nu_k = \mathbb{E} (X)_k = \mathbb{E} X(X-1)\dots(X-k+1)$ ,  $k = 1, \dots$ . Here and below we use the Pochhammer symbol  $(x)_k$  for the falling factorial  $x(x-1)\dots(x-(k-1))$ .

The sequences  $(\mu_k)$  and  $(\kappa_k)$  satisfy  $\mu_k = \kappa_k$  for  $k = 0, 1$  and

$$\kappa_k = \mu_k - \sum_{j=1}^{k-1} \binom{k-1}{j-1} \kappa_j \mu_{k-j} , \quad \text{for } k \geq 2 . \quad (2.2)$$

The sequences  $(\mu_k)$  and  $(\nu_k)$  satisfy  $\mu_0 = \nu_0 = 1$  and

$$\mu_k = \sum_{j=1}^k S_2(k, j) \nu_j , \quad \text{for } k \geq 1 ,$$

where  $S_2(k, j)$  are the Stirling numbers of the second kind defined by the identity  $x^k = \sum_{j=1}^k S_2(k, j)(x)_j$  holding for any  $x \in \mathbb{R}$  (see e.g. [5, (6.10) in Section 6.1]). The validity of (2.2) follows from the relation

$$\phi(t) = \exp(\psi(t)) .$$

Following this probabilistic terminology for any sequence of real numbers  $(a_k)$  one can define its cumulant sequence  $(b_k)$  by

$$\sum a_k \frac{t^k}{k!} = \exp \left\{ \sum b_k \frac{t^k}{k!} \right\} . \quad (2.3)$$

This relation is known in combinatorics under the name exponential formula and its combinatorial interpretation when both  $(a_n)$  and  $(b_n)$  are non-negative integers may be found, for example, in [20, Section 5.1]. Properties of sequences related by (2.2) are usually proved in the combinatorial or probabilistic context, but since the proofs typically involve operating on either side of (2.2) and equating coefficients, they hold

universally. Thus, for example, if  $(a_n)$  and  $(b_n)$  satisfy (2.3) then (see e.g. [20, Proposition 5.1.7] for a proof)

$$b_k = a_k - \sum_{j=1}^{k-1} \binom{k-1}{j-1} b_j a_{k-j}, \quad k \geq 1,$$

where we adopt the usual convention that a sum over an empty range is zero. Note that the sequence  $\underline{b} = (b_k)$  is obtained as a transformation  $f = (f_k)$  of the sequence  $\underline{a} = (a_k)$ , that is  $\underline{b} = f(\underline{a})$ , where  $f_k$ 's are defined recursively by  $f_1(\underline{x}) = x_1$  and

$$f_k(\underline{x}) = x_k - \sum_{j=1}^{k-1} \binom{k-1}{j-1} f_j(\underline{x}) x_{k-j}, \quad k > 1. \quad (2.4)$$

For a future use we record here that if  $\underline{b} = (b_n)$  is a cumulant sequence for a sequence of numbers  $\underline{a} = (a_n)$ , that is,  $\underline{b} = f(\underline{a})$  then for any  $J \geq 1$

$$b_J = \sum_{\pi \subset J} D_\pi \prod_{i=1}^J a_i^{m_i}, \quad \text{where} \quad D_\pi = \frac{(-1)^{\sum_{i=1}^J m_i - 1} J!}{\prod_{i=1}^J (i!)^{m_i} \sum_{i=1}^J m_i} \binom{J}{m_1, \dots, m_J}, \quad (2.5)$$

and where the sum is over all partitions  $\pi$  of a positive integer  $J$ , i.e. over all vectors  $\pi = (m_1, \dots, m_J)$  with nonnegative integer components which satisfy  $\sum_{i=1}^J i m_i = J$  (for a proof, see e.g. [12, Section 3.14]).

Returning for a moment to the probabilistic notions introduced at the beginning of this section we conclude that  $\underline{\kappa} = f(\underline{\mu})$  with  $f$  defined by (2.4).

Similarly for any sequence of real numbers  $(a_k)$  one can define recursively the factorial sequence  $(c_k)$  by

$$a_k = \sum_{j=1}^k S_2(k, j) c_j, \quad \text{for } k \geq 1.$$

Consequently,

$$b_k = \sum_{j=1}^k S_2(k, j) f_j(\underline{c}), \quad k = 1, 2, \dots \quad (2.6)$$

This observation is useful for proving convergence in law to the standard normal variable.

**Proposition 1.** *Let  $(S_n)$  be a sequence of random variables having all moments. Assume that*

$$\mathbb{V}\text{ar} S_n \xrightarrow{n \rightarrow \infty} \infty \quad (2.7)$$

and

$$\frac{\mathbb{E} S_n}{\mathbb{V}\text{ar}^{\frac{3}{2}} S_n} \xrightarrow{n \rightarrow \infty} 0. \quad (2.8)$$

For any  $n = 1, 2, \dots$ , let  $\underline{c}_n = (c_{k,n})_{k=1,\dots}$  be a sequence of factorial moments of  $S_n$ , that is  $c_{k,n} = \mathbb{E}(S_n)_k$ ,  $k = 1, 2, \dots$ , and let  $f_{J,n} = f_J(\underline{c}_n)$  (where  $f_J$  is defined by (2.4)) be the  $J$ th factorial cumulant of  $S_n$ ,  $J = 1, 2, \dots$ . Assume that

$$\frac{f_{J,n}}{\mathbb{Var}^{\frac{J}{2}} S_n} \xrightarrow{n \rightarrow \infty} 0 \quad \text{for any } J \geq 3. \quad (2.9)$$

Then

$$U_n = \frac{S_n - \mathbb{E} S_n}{\sqrt{\mathbb{Var} S_n}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (2.10)$$

*Proof.* We will use the cumulant convergence theorem (see e.g. [8, Theorem 6.14]). Let  $\kappa_{J,n}$  denote the  $J$ th cumulant of  $U_n$  and recall that all cumulants of the standard normal distribution are zero except of the cumulant of order 2 which is 1. It is obvious that  $\kappa_{1,n} = 0$  and  $\kappa_{2,n} = 1$ . Therefore to prove (2.10) it suffices to show that  $\kappa_{J,n} \rightarrow 0$  for any  $J \geq 3$ . By (2.6)

$$\kappa_{J,n} = \frac{\sum_{j=1}^J S_2(J, j) f_{j,n}}{\mathbb{Var}^{\frac{J}{2}} S_n}.$$

Fix arbitrary  $J \geq 3$ . To prove that  $\kappa_{J,n} \rightarrow 0$  it suffices to show that

$$\frac{f_{j,n}}{\mathbb{Var}^{\frac{J}{2}} S_n} \rightarrow 0 \quad \text{for all } j = 1, 2, \dots, J. \quad (2.11)$$

Note first that by (2.4)

$$f_{1,n} = \mathbb{E} S_n \quad \text{and} \quad f_{2,n} = \mathbb{Var}(S_n) - \mathbb{E} S_n$$

Therefore the assumptions (2.7) and (2.8) imply (2.11) for  $j = 1, 2$ .

If  $j \in \{3, \dots, J-1\}$  we write

$$\frac{f_{j,n}}{\mathbb{Var}^{\frac{J}{2}} S_n} = \frac{f_{j,n}}{\mathbb{Var}^{\frac{j}{2}} S_n} \frac{1}{\mathbb{Var}^{\frac{J-j}{2}} S_n}.$$

By (2.9) with  $J = j$  the first factor tends to zero and by (2.7) the second factor tends to zero.

Finally, for  $j = J$  the conditions (2.11) and (2.9) are identical.  $\square$

The above result is particularly useful when the factorial moments of  $S_n$  are available in a nice form. We will now describe a general situation when this happens.

For any numbers  $\delta_j$ ,  $j = 1, \dots, N$ , assuming values 0 or 1 we have

$$x^{\sum_{j=1}^N \delta_j} = \sum_{m=0}^{\sum_{j=1}^N \delta_j} \binom{\sum_{j=1}^N \delta_j}{m} (x-1)^m = 1 + \sum_{m=1}^N (x-1)^m \sum_{1 \leq j_1 < \dots < j_m \leq N} \delta_{j_1} \dots \delta_{j_m}.$$

Therefore if  $(\epsilon_1, \dots, \epsilon_N)$  is a random vector valued in  $\{0, 1\}^N$  and  $S = \sum_{i=1}^N \epsilon_i$  then

$$\mathbb{E} e^{tS} = 1 + \sum_{m=1}^{\infty} (e^t - 1)^m \sum_{1 \leq j_1 < \dots < j_m \leq N} \mathbb{P}(\epsilon_{j_1} = \epsilon_{j_2} = \dots = \epsilon_{j_m} = 1) .$$

Comparing this formula with (2.1) we conclude that factorial moments of  $S$  have the form

$$\mathbb{E}(S)_k = k! \sum_{1 \leq j_1 < \dots < j_k \leq N} \mathbb{P}(\epsilon_{j_1} = \epsilon_{j_2} = \dots = \epsilon_{j_k} = 1) =: c_k , \quad k = 1, 2, \dots \quad (2.12)$$

If, in addition, the random variables  $(\epsilon_1, \dots, \epsilon_N)$  are exchangeable then the above formula simplifies to

$$\mathbb{E}(S)_k = (N)_k \mathbb{P}(\epsilon_1 = \dots = \epsilon_k = 1) =: c_k , \quad k = 1, 2, \dots \quad (2.13)$$

As we will see in Section 5 our sufficient condition for asymptotic normality will work well for several setups falling in such a scheme. This will be preceded by a derivation of new identities for integer partitions which will give a major enhancement of the tools we will use to prove limit theorems.

### 3 Partitions identities

Recall that a partition  $\pi$  of a positive integer  $J$  is any vector  $\pi = (m_1, \dots, m_J)$  with nonnegative integer components which satisfy  $\sum_{i=1}^J i m_i = J$ . As we discussed earlier (see (2.5)) if  $\underline{b} = (b_n) = f(\underline{a})$  is a cumulant sequence of  $\underline{a} = (a_n)$  with  $f$  defined by (2.4) then

$$b_J = \sum_{\pi \subset J} D_\pi \prod_{i=1}^J a_i^{m_i} ,$$

with  $(D_\pi)$  as in (2.5). Note that for  $J \geq 2$

$$\sum_{\pi \subset J} D_\pi = 0 . \quad (3.14)$$

This follows from the fact that all the cumulants of the constant random variable  $X = 1$  a.s. are zero except of the first one.

For  $\pi = (m_1, \dots, m_J)$  we denote  $H_\pi(s) = \sum_{i=1}^J i^s m_i$ ,  $s = 0, 1, 2, \dots$ . The main result of this section is the identity which extends considerably (3.14).

**Proposition 2.** Assume  $J \geq 2$ . Let  $I \geq 1$  and  $s_i \geq 1$ ,  $i = 1, \dots, I$  be such that

$$\sum_{i=1}^I s_i \leq J + I - 2 .$$

Then

$$\sum_{\pi \subset J} D_\pi \prod_{i=1}^I H_\pi(s_i) = 0 . \quad (3.15)$$

*Proof.* We use induction with respect to  $J$ . Note that if  $J = 2$  then  $I$  may be arbitrary but  $s_i = 1$  for all  $i = 1, \dots, I$ . Thus the identity (3.15) follows from (3.14) since  $H_\pi(1) = J$  for any  $J$  and any  $\pi \subset J$ .

Now assume that the result holds true for  $J = 2, \dots, K-1$ , and consider the case of  $J = K$ . That is, we want to study

$$\sum_{\pi \subset K} D_\pi \prod_{i=1}^I H_\pi(s_i)$$

under the condition  $\sum_{i=1}^I s_i \leq K + I - 2$ .

Let us introduce functions  $g_i, i = 1, \dots, K$ , by letting

$$g_i(m_1, \dots, m_K) = (\tilde{m}_1, \dots, \tilde{m}_{K-1}) = \begin{cases} (m_1, \dots, m_{i-1} + 1, m_i - 1, \dots, m_{K-1}) & \text{if } i \neq 1, K, \\ (m_1 - 1, m_2, \dots, m_{K-1}) & \text{if } i = 1, \\ (m_1, \dots, m_{K-2}, m_{K-1} + 1) & \text{if } i = K. \end{cases}$$

Note that  $g_i : \{\pi \subset K : m_i \geq 1\} \rightarrow \{\tilde{\pi} \subset (K-1) : \tilde{m}_{i-1} \geq 1\}, i = 1, \dots, K$ , are bijections. Here for consistency we assume  $\tilde{m}_0 = 1$ .

Observe that for any  $s$ , any  $\pi \subset K$  such that  $m_i \geq 1$ , and for  $\tilde{\pi} = g_i(\pi) \subset (K-1)$  we have

$$H_{\tilde{\pi}}(s) = H_\pi(s) - i^s + (i-1)^s = H_\pi(s) - 1 - A_s(i-1),$$

where  $A_s(i-1) = \sum_{k=1}^{s-1} \binom{s}{k} (i-1)^k$  is a polynomial of the degree  $s-1$  in the variable  $i-1$  with constant term equal to zero. In particular, for  $i = 1$  we have  $H_{\tilde{\pi}}(s) = H_\pi(s) - 1$ . Therefore, expanding  $H_\pi(s_1)$  we obtain

$$\sum_{\pi \subset K} D_\pi \prod_{i=1}^I H_\pi(s_i) = \sum_{i=1}^K i^{s_1} \sum_{\substack{\pi \subset K \\ m_i \geq 1}} m_i D_\pi \prod_{j=2}^I [H_{\tilde{\pi}}(s_j) + 1 + A_{s_j}(i-1)].$$

Note that if  $m_i \geq 1$  then

$$\frac{1}{K!} i m_i D_\pi = \begin{cases} \frac{(-1)^{M_\pi-1} M_\pi!}{M_\pi(m_1-1)! \prod_{k=2}^{K-1} m_k! (k!)^{m_k}} = -\frac{1}{(K-1)!} \sum_{j=1}^{K-1} D_{\tilde{\pi}} \tilde{m}_j & \text{if } i = 1, \\ \frac{(-1)^{M_\pi-1} M_\pi!}{M_\pi(m_i-1)! (i-1)! (i!)^{m_i-1} \prod_{\substack{k=2 \\ k \neq i}}^K m_k! (k!)^{m_k}} = \frac{1}{(K-1)!} D_{\tilde{\pi}} \tilde{m}_{i-1} & \text{if } i = 2, \dots, K, \end{cases}$$

where  $\tilde{\pi} = (\tilde{m}_1, \dots, \tilde{m}_{K-1}) = g_i(\pi)$ , respectively. Therefore,

$$\begin{aligned} \frac{1}{K} \sum_{\pi \subset K} D_\pi \prod_{i=1}^I H_\pi(s_i) &= - \sum_{i=1}^{K-1} \sum_{\tilde{\pi} \subset (K-1)} D_{\tilde{\pi}} \tilde{m}_i \prod_{j=2}^I [H_{\tilde{\pi}}(s_j) + 1] \\ &+ \sum_{i=2}^K \sum_{\substack{\tilde{\pi} \subset (K-1) \\ \tilde{m}_{i-1} \geq 1}} D_{\tilde{\pi}} \tilde{m}_{i-1} i^{s_1-1} \prod_{j=2}^I [H_{\tilde{\pi}}(s_j) + 1 + A_{s_j}(i-1)]. \end{aligned} \quad (3.16)$$

The second term in the above expression can be written as

$$\sum_{i=1}^{K-1} \sum_{\tilde{\pi} \subset (K-1)} D_{\tilde{\pi}} \tilde{m}_i (i+1)^{s_1-1} \prod_{j=2}^I [H_{\tilde{\pi}}(s_j) + 1 + A_{s_j}(i)] . \quad (3.17)$$

Note that

$$\begin{aligned} & (i+1)^{s_1-1} \prod_{j=2}^I [H_{\tilde{\pi}}(s_j) + 1 + A_{s_j}(i)] \\ &= \sum_{r=0}^{s_1-1} \binom{s_1-1}{r} \sum_{M=0}^{I-1} \sum_{2 \leq u_1 < \dots < u_M \leq I} i^r \prod_{h=1}^M (H_{\tilde{\pi}}(s_{u_h}) + 1) \prod_{\substack{2 \leq j \leq I \\ j \notin \{u_1, \dots, u_M\}}} A_{s_j}(i) . \end{aligned}$$

The term with  $r = 0$  and  $M = I - 1$  in the above expression is  $\prod_{j=2}^I [H_{\tilde{\pi}}(s_j) + 1]$ , thus this term of the sum (3.17) cancels with the first term of (3.16).

Hence, we need only to show that for any  $r \in \{1, \dots, s_1 - 1\}$ , any  $M \in \{0, \dots, I - 1\}$ , and any  $2 \leq u_1 < \dots < u_M \leq I$

$$\sum_{\tilde{\pi} \subset (K-1)} D_{\tilde{\pi}} \prod_{h=1}^M (H_{\tilde{\pi}}(s_{u_h}) + 1) \sum_{i=1}^{K-1} \tilde{m}_i i^r \prod_{\substack{2 \leq j \leq I \\ j \notin \{u_1, \dots, u_M\}}} A_{s_j}(i) = 0. \quad (3.18)$$

Observe that the expression

$$\sum_{i=1}^{K-1} \tilde{m}_i i^r \prod_{\substack{2 \leq j \leq I \\ j \notin \{u_1, \dots, u_M\}}} A_{s_j}(i)$$

is a linear combination of  $H_{\tilde{\pi}}$  functions with the largest value of an argument equal to

$$\sum_{\substack{2 \leq j \leq I \\ j \notin \{u_1, \dots, u_M\}}} (s_j - 1) + r .$$

Therefore the left-hand side of (3.18) is a respective linear combination of terms of the form

$$\sum_{\tilde{\pi} \subset (K-1)} D_{\tilde{\pi}} \prod_{w=1}^W H_{\tilde{\pi}}(t_w) \quad (3.19)$$

where

$$\sum_{w=1}^W t_w \leq \sum_{j=1}^M s_{u_j} - (M - W + 1) + \sum_{\substack{2 \leq j \leq I \\ j \notin \{u_1, \dots, u_M\}}} (s_j - 1) + s_1 - 1 \leq \sum_{i=1}^I s_i - (M - W + 1) - (I - 1 - M) - 1 .$$

But we assumed that  $\sum_{i=1}^I s_i \leq K + I - 2$ . Therefore

$$\sum_{w=1}^W t_w \leq K + I - 2 - (M - W + 1) - (I - 1 - M) - 1 = (K - 1) + W - 2 .$$



Now, by the inductive assumption it follows that any term of the form (3.19) is zero, thus proving (3.18).  $\square$

Note that in a similar way one can prove that (3.15) is no longer true when

$$\sum_{i=1}^I s_i = J + I - 1 .$$

**Remark** Richard Stanley [21] provided the following combinatorial description of the left hand side of (3.15): put

$$F_n(x) = \sum_k S_2(n, k) x^k ,$$

and let  $(s_i)$  be a sequence of positive integers. Then the left hand side of (3.15) is a coefficient of  $x^J / J!$  in

$$\sum_{\mathcal{P}} (-1)^{|\mathcal{P}|-1} (|\mathcal{P}| - 1)! \prod_{B \in \mathcal{P}} F_{\sigma B}(x) ,$$

where the sum ranges over all partitions  $\mathcal{P}$  of a set  $\{1, \dots, I\}$  into  $|\mathcal{P}|$  nonempty pairwise disjoint subsets, and where for any such subset  $B \in \mathcal{P}$

$$\sigma B = \sum_{i \in B} s_i .$$

In the simplest case when  $I = 1$  for any positive integer  $s_1$ , the left-hand side of equation (3.15) is equal to  $J! S_2(s_1, J)$ . Since this is the number of surjective maps from an  $s_1$ -element set to a  $J$ -element set, it must be 0 for  $s_1 < J$  which is exactly what Proposition 2 asserts. It is not clear to us how easy it would be to show that (3.15) holds for the larger values of  $I$ .

## 4 CLT - general setup

To illustrate and motivate how our approach is intended to work, consider a sequence  $(S_n)$  of Poisson random variables, where  $S_n \sim \text{Poisson}(\lambda_n)$ . As is well-known if  $\lambda_n \rightarrow \infty$  then  $U_n = \frac{S_n - \mathbb{E} S_n}{\sqrt{\text{Var} S_n}}$  converges in distribution to  $\mathcal{N}(0, 1)$ . To see how it follows from our approach, note that  $\mathbb{E} S_n = \text{Var} S_n = \lambda_n$  and therefore the assumptions (2.7) and (2.8) of Proposition 1 are trivially satisfied. Moreover the factorial moments of  $S_n$  are  $c_{i,n} = \lambda_n^i$ . Consequently,  $\prod_{i=1}^J c_{i,n}^{m_i} = \lambda_n^J$  for any partition  $\pi = (m_1, \dots, m_J)$  of  $J \in \mathbb{N}$  and thus

$$f_{J,n} = f_J(\mathcal{L}_n) = \sum_{\pi \subset J} D_\pi \prod_{i=1}^J c_{i,n}^{m_i} = \lambda_n^J \sum_{\pi \subset J} D_\pi .$$

It now follows from the simplest case (3.14) of our partition identity that  $f_{J,n} = 0$  as long as  $J \geq 2$ . Hence the assumption (2.9) of Proposition 1 is also trivially satisfied and we conclude the asymptotic normality of  $(U_n)$ .

The key in the above argument was, of course, a very simple form of the factorial moments  $c_{i,n}$  of  $S_n$  which resulted in factorization of the products of  $c_{i,n}^{m_i}$  in the expression for  $f_{J,n}$ . It is not unreasonable,

however, to expect that if the expression for moments does not depart too much from the form it took for Poisson variable, then with the full strength of Proposition 2 one might be able to prove the clt. This is the essence of condition (4.20) in the next result. This condition when combined with the extension of (3.14) given in Proposition 2, allows to refine considerably Proposition 1 towards a possible use in schemes of summation of indicators, as was suggested in the final part of Section 2.

**Theorem 3.** *Let  $(S_n)$  be a sequence of random variables with factorial moments  $c_{i,n}$ ,  $i, n = 1, 2, \dots$ . Assume that (2.7) and (2.8) are satisfied and suppose that  $c_{i,n}$  can be decomposed as*

$$c_{i,n} = L_n^i \exp \left( \sum_{j \geq 1} \frac{Q_{j+1}^{(n)}(i)}{jn^j} \right), \quad i, n = 1, 2, \dots, \quad (4.20)$$

where  $(L_n)$  is a sequence of real numbers and  $Q_j^{(n)}$  is a polynomial of degree at most  $j$  such that

$$|Q_j^{(n)}(i)| \leq (Ci)^j \quad \forall i \in \mathbb{N} \quad (4.21)$$

with  $C > 0$  a constant not depending on  $n$  or  $j$ . Assume further that for all  $J \geq 3$

$$\frac{L_n^J}{n^{J-1} \mathbb{V} \text{ar}^{\frac{J}{2}} S_n} \xrightarrow{n \rightarrow \infty} 0. \quad (4.22)$$

Then

$$U_n = \frac{S_n - \mathbb{E} S_n}{\sqrt{\mathbb{V} \text{ar} S_n}} \xrightarrow{d} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty. \quad (4.23)$$

*Proof.* Due to Proposition 1 we need only to show that (2.9) holds. The representation (4.20) implies

$$f_{J,n} = \sum_{\pi \subset J} D_\pi \prod_{i=1}^J c_{i,n}^{m_i} = L_n^J \sum_{\pi \subset J} D_\pi e^{z_\pi(J,n)},$$

where

$$z_\pi(J,n) = \sum_{j \geq 1} \frac{A_\pi^{(n)}(j)}{jn^j} \quad \text{with} \quad A_\pi^{(n)}(j) = \sum_{i=1}^J m_i Q_{j+1}^{(n)}(i).$$

Fix arbitrary  $J \geq 3$ . To prove (2.9), in view of (4.22) it suffices to show that  $\sum_{\pi \subset J} D_\pi e^{z_\pi(J,n)}$  is of order  $n^{-(J-1)}$ . To do that, we expand  $e^{z_\pi(J,n)}$  into power series to obtain

$$\begin{aligned} \sum_{\pi \subset J} D_\pi e^{z_\pi(J,n)} &= \sum_{\pi \subset J} D_\pi e^{\sum_{j \geq 1} \frac{1}{jn^j} A_\pi^{(n)}(j)} = \sum_{\pi \subset J} D_\pi \sum_{s=0}^{\infty} \frac{1}{s!} \left( \sum_{j \geq 1} \frac{1}{jn^j} A_\pi^{(n)}(j) \right)^s \\ &= \sum_{s \geq 1} \frac{1}{s!} \sum_{l \geq s} \frac{1}{n^l} \sum_{\substack{j_1, \dots, j_s \geq 1 \\ \sum_{k=1}^s j_k = l}} \frac{1}{\prod_{k=1}^s j_k} \sum_{\pi \subset J} D_\pi \prod_{k=1}^s A_\pi^{(n)}(j_k). \end{aligned}$$

We claim that whenever  $\sum_{k=1}^s j_k \leq J - 2$  then

$$\sum_{\pi \subset J} D_\pi \prod_{k=1}^s A_\pi^{(n)}(j_k) = 0. \quad (4.24)$$

To see this, note that by changing the order of summation in the expression for  $A_\pi^{(n)}(j)$  we can write it as

$$A_\pi^{(n)}(j) = \sum_{k=0}^{j+1} \alpha_{k,j+1}^{(n)} H_\pi(k),$$

where  $(\alpha_{k,j}^{(n)})$  are the coefficients of the polynomial  $Q_j^{(n)}$ , that is  $Q_j^{(n)}(x) = \sum_{k=0}^j \alpha_{k,j}^{(n)} x^k$ . Consequently, (4.24) follows from identity (3.15).

To handle the terms for which  $\sum_{k=1}^s j_k > J - 2$  note that

$$|A_\pi^{(n)}(j)| \leq (CJ)^{j+1} \sum_{i=1}^J m_i < K (CJ)^j,$$

where  $K > 0$  is a constant depending only on  $J$  (and not on the partition  $\pi$ ). Hence,

$$\left| \sum_{\pi \subset J} D_\pi \prod_{k=1}^s A_\pi^{(n)}(j_k) \right| \leq \sum_{\pi \subset J} |D_\pi| \prod_{k=1}^s K (CJ)^{j_k} \leq \tilde{C} K^s (CJ)^{\sum_{k=1}^s j_k}, \quad (4.25)$$

where  $\tilde{C} = \sum_{\pi \subset J} |D_\pi|$  is a constant depending only on  $J$ . Therefore, restricting the sum according to (4.24) and using (4.25) we get

$$\begin{aligned} \left| \sum_{\pi \subset J} D_\pi e^{z_\pi(J,n)} \right| &\leq \sum_{s \geq 1} \frac{1}{s!} \sum_{l \geq \max\{s, J-1\}} \frac{1}{n^l} \sum_{\substack{j_1, \dots, j_s \geq 1 \\ \sum_{k=1}^s j_k = l}} \frac{1}{\prod_{k=1}^s j_k} \left| \sum_{\pi \subset J} D_\pi \prod_{k=1}^s A_\pi^{(n)}(j_k) \right| \\ &\leq \tilde{C} \sum_{s \geq 1} \frac{K^s}{s!} \sum_{l \geq J-1} \frac{1}{n^l} l^s (CJ)^l. \end{aligned}$$

Here we used the inequality

$$\sum_{\substack{j_1, \dots, j_s \geq 1 \\ \sum_{k=1}^s j_k = l}} \frac{1}{\prod_{k=1}^s j_k} < l^s,$$

which may be seen by trivially bounding the sum by the number of its terms. Now we change the order of summations arriving at

$$\left| \sum_{\pi \subset J} D_\pi e^{z_\pi(J,n)} \right| \leq \tilde{C} \sum_{l \geq J-1} \left( \frac{CJ}{n} \right)^l \sum_{s \geq 1} \frac{(lK)^s}{s!} \leq \tilde{C} \sum_{l \geq J-1} \left( \frac{CJe^K}{n} \right)^l = \tilde{C} \left( \frac{CJe^K}{n} \right)^{J-1} \sum_{l \geq 0} \left( \frac{CJe^K}{n} \right)^l.$$

The result follows since for  $n$  sufficiently large (such that  $CJe^K < n$ ) the series in the last expression converges.  $\square$

**Remark 4.** A typical way Theorem 3 will be applied is as follows. Assume that  $\mathbb{E} S_n$  and  $\mathbb{V}ar S_n$  are of the same order  $n$ . Then obviously, (2.7) and (2.8) are satisfied. Assume also that (4.20) and (4.21) hold and that  $L_n$  is also of order  $n$ . Then clearly (4.22) is satisfied and thus (4.23) holds true.

## 5 Applications

In this section we show how the tools developed in previous section and in particular, the decomposition (4.20) together with the condition (4.22) can be conveniently used for proving clt's in several situations, mostly in summation schemes of 0 – 1 random variables, as was indicated in Section 2. First, four more or less standard limit results for the binomial, negative binomial, hypergeometric and negative hypergeometric schemes will be reproved. Then more involved schemes of allocation problems for distinguishable, indistinguishable balls, colored balls, and random forests will be considered. The clt's for the number of boxes with exactly  $r$  balls in the first two problems and for the number of trees with exactly  $r$  non-root vertices in the third problem will be derived. While the clt in the case of distinguishable balls has been known in the literature for years (see for instance [16]), the main focus in the other two cases appears to be on the local limit theorems (see, e.g. [14, 15, 19]). We have not found any references for the asymptotic normality results for the problems we consider in gas models.

The models 5.2.1–5.2.4 are examples of the generalized allocation scheme (see [14, 15] for more details), that is

$$\mathbb{P}(\xi_1^{(n)} = k_1, \dots, \xi_N^{(n)} = k_N) = \mathbb{P}(\eta_1 = k_1, \dots, \eta_N = k_N | \eta_1 + \dots + \eta_N = n), \quad (5.26)$$

while models 5.3.1 and 5.3.2 are examples of what may be called the generalized inverse allocation scheme, that is

$$\mathbb{P}(\xi_1^{(n)} = k_1, \dots, \xi_N^{(n)} = k_N) = \mathbb{P}(\eta_1 = k_1, \dots, \eta_N = k_N | \eta_0 + \eta_1 + \dots + \eta_N = n + k_1 + \dots + k_N), \quad (5.27)$$

where  $\eta_0, \eta_1, \dots, \eta_N$  are independent random variables. In practical situations of gas models  $(\eta_j)$  are identically distributed, and in the case of gas we assume that  $\eta_j$ 's have the same distribution for  $j = 1, \dots, N$  which may differ from the distribution of  $\eta_0$ .

In the derivations below we will often use the following expansion

$$\left(1 - \frac{a}{b}\right)^c = e^{c \log(1 - \frac{a}{b})} = e^{-c \sum_{j=1}^{\infty} \frac{a^j}{j b^j}}, \quad (5.28)$$

which is valid for any  $0 < |a| < |b|$  and any real  $c$ . We also recall (see e.g. [5, Chapter 6.5]) that

$$\mathcal{Q}_{j+1}(M) := \sum_{k=1}^{M-1} k^j = \frac{1}{j+1} \sum_{l=1}^{j+1} \binom{j+1}{l} B_{j+1-l} M^l, \quad (5.29)$$

where  $(B_k)$  are the Bernoulli numbers. Clearly,  $\mathcal{Q}_j$  is a polynomial of degree  $j$  satisfying (4.21) with  $C = 1$ . For notational convenience we let

$$T(m, t) = \prod_{k=1}^{m-1} \left(1 - \frac{k}{t}\right)$$

for  $t > 0$  and integer  $m > 0$ . It follows from (5.28) and (5.29) that for  $t > m$

$$T(m, t) = e^{-\sum_{j \geq 1} \frac{1}{j t^j} \mathcal{Q}_{j+1}(m)}. \quad (5.30)$$

## 5.1 Classical discrete distributions

In this subsection we re-derive asymptotic normality of

$$\frac{S_n - \mathbb{E} S_n}{\sqrt{\mathbb{V}\text{ar } S_n}}$$

for laws of  $S_n$  belonging to four classical families of discrete distributions: binomial, negative binomial, hypergeometric and negative hypergeometric.

### 5.1.1 Binomial scheme

Let  $(\epsilon_i)$  be a sequence of iid Bernoulli random variables,  $P(\epsilon_1 = 1) = p = 1 - P(\epsilon_1 = 0)$ . Then  $S_n = \sum_{i=1}^n \epsilon_i$  has the binomial  $b(n, p)$  distribution. To see how Theorem 3 allows in a simple way to re-derive de Moivre-Laplace theorem

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} \mathcal{N}(0, 1) \quad (5.31)$$

we set  $L_n = np$ . Then  $\mathbb{E} S_n = L_n$  and  $\mathbb{V}\text{ar } S_n = L_n(1-p)$ . Furthermore,  $\mathbb{P}(\epsilon_1 = \dots = \epsilon_i = 1) = p^i$  and thus by (2.13) it follows that the  $i$ th factorial moment of  $S_n$  is

$$c_{i,n} = \mathbb{E}(S_n)_i = (n)_i p^i = L_n^i T(i, n) . \quad (5.32)$$

Thus (5.30) implies representation (4.20) with  $Q_{j+1} = -Q_{j+1}$  and (5.31) follows from Remark 4.

### 5.1.2 Negative binomial scheme

Let  $S_n$  denote the number of failures until the  $n$ th success in Bernoulli trials with  $p$  being the probability of a success in a single trial, that is,  $S_n$  is negative binomial  $nb(n, p)$  with

$$\mathbb{P}(S_n = k) = \binom{n+k-1}{k} (1-p)^k p^n \quad k = 0, 1, \dots$$

We will show how Theorem 3 allows in a simple way to re-derive the clt for  $S_n$ , which says that for  $n \rightarrow \infty$

$$\frac{pS_n - n(1-p)}{\sqrt{n(1-p)}} \xrightarrow{d} \mathcal{N}(0, 1) . \quad (5.33)$$

Set  $L_n = n \frac{1-p}{p}$  so that  $\mathbb{E} S_n = L_n$  and  $\mathbb{V}\text{ar } S_n = \frac{L_n}{p}$ . Furthermore, the  $i$ th factorial moment of  $S_n$  is easily derived as

$$c_{i,n} = \mathbb{E}(S_n)_i = L_n^i T(i, -n) .$$

Hence (4.20) holds with  $Q_{j+1} = (-1)^{j+1} Q_{j+1}$  and thus (5.33) follows from Remark 4.

### 5.1.3 Hypergeometric scheme

From an urn containing  $N$  white and  $M$  black balls we draw subsequently without replacement  $n$  balls ( $n \leq \min\{M, N\}$ ). For  $i = 1, \dots, n$ , let  $\epsilon_i = 1$  if a white ball is drawn at the  $i$ th drawing and let  $\epsilon_i = 0$  otherwise. Then  $S_n = \sum_{i=1}^n \epsilon_i$  has a hypergeometric distribution  $Hg(N, M; n)$ , that is,

$$\mathbb{P}(S_n = k) = \frac{\binom{N}{k} \binom{M}{n-k}}{\binom{N+M}{n}}, \quad k = 0, 1, \dots, n.$$

Using again Theorem 3 we will show that under the assumptions  $N = N(n) \rightarrow \infty$ ,  $M = M(n) \rightarrow \infty$ , and  $\frac{N}{N+M} \rightarrow p \in (0, 1)$  with  $n \rightarrow \infty$

$$\frac{(N+M)S_n - nN}{\sqrt{nNM(N+M-n)/(N+M-1)}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (5.34)$$

Setting  $L_n = \frac{nN}{N+M}$  we have

$$\mathbb{E} S_n = L_n \quad \text{and} \quad \mathbb{V}ar S_n = L_n \frac{M(N+M-n)}{(N+M)(N+M-1)}. \quad (5.35)$$

Moreover, on noting that  $(\epsilon_1, \dots, \epsilon_n)$  is exchangeable by (2.13) we get

$$c_{i,n} = \mathbb{E} (S_n)_i = (n)_i \mathbb{P}(\epsilon_1 = \dots = \epsilon_i = 1) = \frac{(n)_i (N)_i}{(N+M)_i} = L_n^i \frac{T(i, n) T(i, N)}{T(i, N+M)}.$$

Similarly as in earlier schemes we obtain representation (4.20) with

$$\mathcal{Q}_{j+1} = \left[ -1 - \left( \frac{n}{N} \right)^j + \left( \frac{n}{N+M} \right)^j \right] \mathcal{Q}_{j+1}.$$

Moreover, the condition (4.22) is fulfilled since  $\mathbb{E} S_n$ ,  $\mathbb{V}ar S_n$  and  $L_n$  are all of order  $n$ . See again Remark 4 to conclude that (5.34) holds true.

### 5.1.4 Negative hypergeometric scheme

Let  $S_n$  be a random variable with negative hypergeometric distribution of the first kind, that is,

$$\mathbb{P}(S_n = k) = \binom{n}{k} \frac{B(\alpha_n + k, \beta_n + n - k)}{B(\alpha_n, \beta_n)} \quad k = 0, 1, \dots, n,$$

with  $\alpha_n = n\alpha$  and  $\beta_n = n\beta$ . The clt for  $S_n$ , says that for  $n \rightarrow \infty$

$$\frac{(\alpha + \beta)^{\frac{3}{2}} S_n - n\alpha\sqrt{\alpha + \beta}}{\sqrt{n\alpha\beta(1 + \alpha + \beta)}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (5.36)$$

To quickly derive it from Theorem 3 let  $L_n = \frac{n\alpha}{\alpha+\beta}$  and note that  $\mathbb{E} S_n = L_n$  and  $\mathbb{V}ar S_n = L_n \frac{n\beta(1+\alpha+\beta)}{(\alpha+\beta)^2(n\alpha+n\beta+1)}$ . Further, the  $i$ th factorial moment of  $S_n$  is easily derived as

$$c_{i,n} = \mathbb{E}(S_n)_i = L_n^i \frac{T(i, n) T(i, -\alpha n)}{T(i, -(\alpha + \beta)n)}.$$

Thus again due to (5.30) we conclude that representation (4.20) holds with

$$Q_{j+1} = \left( -1 - \frac{(-1)^j}{\alpha^j} + \frac{(-1)^j}{(\alpha + \beta)^j} \right) Q_{j+1}(i).$$

The final result follows by Remark 4.

## 5.2 Asymptotics of occupancy in generalized allocation schemes (gas)

In this subsection we will derive asymptotics for  $S_n^{(r)} = \sum_{i=1}^N I(\xi_i^{(n)} = r)$  in several generalized allocation schemes as defined at the beginning of Section 5. As we will see when  $n \rightarrow \infty$  and  $N/n \rightarrow \lambda \in (0, \infty]$  the order of  $\mathbb{E} S_n^{(r)}$  is  $n^r/N^{r-1}$  for any  $r = 0, 1, \dots$ , and the order of  $\mathbb{V}ar S_n^{(r)}$  is the same for  $r \geq 2$ . When  $\lambda = \infty$  and  $r = 0$  or  $1$  the order of  $\mathbb{V}ar S_n^{(r)}$  is  $n^2/N$ . Consequently, we will derive asymptotic normality of

$$\frac{S_n - \mathbb{E} S_n^{(r)}}{\sqrt{n^r/N^{r-1}}}$$

when either:

- (a)  $r \geq 0$  and  $\lambda < \infty$  or,
- (b)  $r \geq 2$ ,  $\lambda = \infty$  and  $\frac{n^r}{N^{r-1}} \rightarrow \infty$ ,

and asymptotic normality of

$$\sqrt{N} \frac{S_n^{(r)} - \mathbb{E} S_n^{(r)}}{n}$$

when  $\lambda = \infty$ ,  $\frac{n^2}{N} \rightarrow \infty$  and  $r = 0, 1$ .

Though in all the cases results read literally the same (with different asymptotic expectations and variances and having different proofs) for the sake of precision we decided to repeat formulations of theorems in each of subsequent cases.

### 5.2.1 Indistinguishable balls

Consider a scheme of a random distribution of  $n$  indistinguishable balls into  $N$  distinguishable boxes, such that all distributions are equiprobable. That is, if  $\xi_i = \xi_i^{(n)}$  denotes the number of balls which fall into the  $i$ th box,  $i = 1, \dots, N$ , then

$$\mathbb{P}(\xi_1 = i_1, \dots, \xi_N = i_N) = \binom{n + N - 1}{n}^{-1}$$

for any  $i_k \geq 0, k = 1, \dots, N$ , such that  $i_1 + \dots + i_N = n$ . Note that this is gas and that (5.26) holds with  $\eta_i \sim \text{Geom}(p), 0 < p < 1$ .

Let  $S_n^{(r)} = \sum_{i=1}^N I(\xi_i = r)$  denote the number of boxes with exactly  $r$  balls. Note that the distribution of  $(\xi_1, \dots, \xi_N)$  is exchangeable. Moreover,

$$\mathbb{P}(\xi_1 = \dots = \xi_i = r) = \frac{\binom{n-ri+N-i-1}{n-ri}}{\binom{n+N-1}{n}}.$$

Therefore by (2.13) we get

$$c_{i,n} = \mathbb{E}(S_n^{(r)})_i = \frac{(N)_i(N-1)_i(n)_{ir}}{(n+N-1)_{i(r+1)}}. \quad (5.37)$$

Consequently,

$$\mathbb{E} S_n^{(r)} = c_{1,n} = \frac{N(N-1)(n)_r}{(n+N-1)_{r+1}}$$

and since  $\mathbb{V}\text{ar} S_n^{(r)} = c_{2,n} - c_{1,n}^2 + c_{1,n}$  we have

$$\mathbb{V}\text{ar} S_n^{(r)} = \frac{N(N-1)^2(N-2)(n)_{2r}}{(n+N-1)_{2r+2}} - \left( \frac{N(N-1)(n)_r}{(n+N-1)_{r+1}} \right)^2 + \frac{N(N-1)(n)_r}{(n+N-1)_{r+1}}.$$

In the asymptotics below we consider the situation when  $n \rightarrow \infty$  and  $\frac{N}{n} \rightarrow \lambda \in (0, \infty]$ . Then for any integer  $r \geq 0$

$$\frac{N^{r-1}}{n^r} \mathbb{E} S_n^{(r)} \rightarrow \left( \frac{\lambda}{1+\lambda} \right)^{r+1} \quad (= 1 \text{ for } \lambda = \infty). \quad (5.38)$$

It is also elementary but more laborious to prove that for any  $r \geq 2$  and  $\lambda \in (0, \infty]$  or  $r = 0, 1$  and  $\lambda \in (0, \infty)$

$$\frac{N^{r-1}}{n^r} \mathbb{V}\text{ar} S_n^{(r)} \rightarrow \left( \frac{\lambda}{1+\lambda} \right)^{r+1} \left( 1 - \frac{\lambda(1+\lambda+(\lambda r-1)^2)}{(1+\lambda)^{r+2}} \right) =: \sigma_r^2 \quad (= 1 \text{ for } \lambda = \infty). \quad (5.39)$$

Additionally, for  $\lambda = \infty$

$$\frac{N}{n^2} \mathbb{V}\text{ar} S_n^{(0)} \rightarrow 1 =: \tilde{\sigma}_0^2 \quad \text{and} \quad \frac{N}{n^2} \mathbb{V}\text{ar} S_n^{(1)} \rightarrow 4 =: \tilde{\sigma}_1^2.$$

Similarly, in this case

$$\frac{N}{n^2} \mathbb{C}\text{ov}(S_n^{(0)}, S_n^{(1)}) \rightarrow -2 \quad \text{and} \quad \frac{N}{n^2} \mathbb{C}\text{ov}(S_n^{(0)}, S_n^{(2)}) \rightarrow 1.$$

and thus for the correlation coefficient we have

$$\rho(S_n^{(0)}, S_n^{(1)}) \rightarrow -1 \quad \text{and} \quad \rho(S_n^{(0)}, S_n^{(2)}) \rightarrow 1. \quad (5.40)$$

Now we are ready to deal with clt's.



**Theorem 5.** Let  $N/n \rightarrow \lambda \in (0, \infty]$ . Let either

(a)  $r \geq 0$  and  $\lambda < \infty$

or

(b)  $r \geq 2$ ,  $\lambda = \infty$  and  $\frac{n^r}{N^{r-1}} \rightarrow \infty$ .

Then

$$\frac{S_n^{(r)} - \mathbb{E} S_n^{(r)}}{\sqrt{n^r/N^{r-1}}} \xrightarrow{d} \mathcal{N}(0, \sigma_r^2) .$$

*Proof.* Note that (5.37) can be rewritten as

$$c_{i,n} = L_n^i \frac{T(i, N) T(i, N-1) T(ir, n)}{T(i(r+1), n+N-1)} \quad \text{with} \quad L_n = \frac{N(N-1)n^r}{(n+N-1)^{r+1}} .$$

Therefore, similarly as in the previous cases, using (5.30) we conclude that representation (4.20) holds with

$$Q_{j+1}(i) = - \left[ \left( \frac{n}{N} \right)^j + \left( \frac{n}{N-1} \right)^j \right] Q_{j+1}(i) - Q_{j+1}(ri) + \left( \frac{n}{n+N-1} \right)^j Q_{j+1}((r+1)i) .$$

To conclude the proof we note that  $\mathbb{E} S_n^{(r)}$ ,  $\mathbb{V}\text{ar} S_n^{(r)}$  and  $L_n$  are of the same order  $n^r/N^{r-1}$  and use Remark 6 stated below.  $\square$

**Remark 6.** If  $\mathbb{E} S_n^{(r)}$  and  $\mathbb{V}\text{ar} S_n^{(r)}$  are of the same order and diverge to  $\infty$  then (2.7) and (2.8) hold. Moreover, if  $L_n$  and  $\mathbb{V}\text{ar} S_n^{(r)}$  are both of order  $n^r/N^{r-1}$  then the left hand side of (4.22) is of order

$$\frac{1}{n^{J-1}} \left( \frac{n^r}{N^{r-1}} \right)^{\frac{J}{2}} = \left( \frac{n}{N} \right)^{\frac{J}{2}(r-1)} \frac{1}{n^{\frac{J}{2}-1}} .$$

That is, when either  $\lambda \in (0, \infty)$  and  $r = 0, 1, \dots$  or  $\lambda = \infty$  and  $r = 2, 3, \dots$  the condition (4.22) is satisfied.

In the remaining cases we use asymptotic correlations

**Theorem 7.** Let  $N/n \rightarrow \infty$  and  $n^2/N \rightarrow \infty$ . Then for  $r = 0, 1$

$$\sqrt{N} \frac{S_n^{(r)} - \mathbb{E} S_n^{(r)}}{n} \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}_r^2) .$$

*Proof.* Due to the second equation in (5.40) it follows that

$$\sqrt{N} \frac{S_n^{(0)} - \mathbb{E} S_n^{(0)}}{\tilde{\sigma}_0 n} - \sqrt{N} \frac{S_n^{(2)} - \mathbb{E} S_n^{(2)}}{\sigma_2 n} \xrightarrow{L^2} 0 .$$

Therefore the result for  $r = 0$  holds. Similarly, for  $r = 1$  it suffices to observe that the first equation in (5.40) implies

$$\sqrt{N} \frac{S_n^{(0)} - \mathbb{E} S_n^{(0)}}{\tilde{\sigma}_0 n} + \sqrt{N} \frac{S_n^{(1)} - \mathbb{E} S_n^{(1)}}{\tilde{\sigma}_1 n} \xrightarrow{L^2} 0 .$$

$\square$

### 5.2.2 Distinguishable balls

Consider a scheme of a random distribution of  $n$  distinguishable balls into  $N$  distinguishable boxes, such that any such distribution is equally likely. Then, if  $\xi_i = \xi_i^{(n)}$  denotes the number of balls which fall into the  $i$ th box,  $i = 1, \dots, N$ ,

$$\mathbb{P}(\xi_1 = i_1, \dots, \xi_N = i_N) = \frac{n!}{i_1! \dots i_N!} N^{-n}$$

for any  $i_l \geq 0$ ,  $l = 1, \dots, N$ , such that  $i_1 + \dots + i_N = n$ . This is a gas with  $\eta_i \sim \text{Poisson}(\lambda)$ ,  $\lambda > 0$ , in (5.26).

For a fixed nonnegative integer  $r$  let

$$S_n^{(r)} = \sum_{i=1}^N I(\xi_i = r)$$

be the number of boxes with exactly  $r$  balls. Obviously, the distribution of  $(\xi_1, \dots, \xi_N)$  is exchangeable and

$$\mathbb{P}(\xi_1 = \dots = \xi_i = r) = \frac{n!}{(r!)^i (n - ir)!} N^{-ir} \left(1 - \frac{i}{N}\right)^{n-ir}.$$

Therefore by (2.13) we get

$$c_{i,n} = \mathbb{E}(S_n^{(r)})_i = \frac{(N)_i (n)_{ir}}{(r!)^i N^{ri}} \left(1 - \frac{i}{N}\right)^{n-ir}. \quad (5.41)$$

Consequently, for any  $r = 0, 1, \dots$

$$\mathbb{E} S_n^{(r)} = c_{1,n} = \frac{(n)_r \left(1 - \frac{1}{N}\right)^{n-r}}{r! N^{r-1}}$$

and

$$\mathbb{V}ar S_n^{(r)} = c_{2,n} - c_{1,n}^2 + c_{1,n} = \frac{(N-1)(n)_{2r} \left(1 - \frac{2}{N}\right)^{n-2r}}{(r!)^2 N^{2r-1}} - \frac{(n)_r^2 \left(1 - \frac{1}{N}\right)^{2(n-r)}}{(r!)^2 N^{2(r-1)}} + \frac{(n)_r \left(1 - \frac{1}{N}\right)^{n-r}}{r! N^{r-1}}.$$

In the asymptotics below we consider the situation when  $n \rightarrow \infty$  and  $\frac{N}{n} \rightarrow \lambda \in (0, \infty]$ . Then for any integer  $r \geq 0$

$$\lim_{n \rightarrow \infty} \frac{N^{r-1}}{n^r} \mathbb{E} S_n^{(r)} = \frac{1}{r!} e^{-\frac{1}{\lambda}} \quad \left( = \frac{1}{r!} \text{ for } \lambda = \infty \right). \quad (5.42)$$

It is also elementary but more laborious to check that for any fixed  $r \geq 2$  and  $\lambda \in (0, \infty]$  or  $r = 0, 1$  and  $\lambda \in (0, \infty)$

$$\lim_{n \rightarrow \infty} \frac{N^{r-1}}{n^r} \mathbb{V}ar S_n^{(r)} = \frac{e^{-\frac{1}{\lambda}}}{r!} \left( 1 - \frac{e^{-\frac{1}{\lambda}} (\lambda + (\lambda r - 1)^2)}{r! \lambda^{r+1}} \right) := \sigma_r^2 \quad \left( = \frac{1}{r!} \text{ for } \lambda = \infty \right). \quad (5.43)$$

Additionally, for  $\lambda = \infty$

$$\frac{N}{n^2} \mathbb{V}ar S_n^{(0)} \rightarrow \frac{1}{2} =: \tilde{\sigma}_0^2 \quad \text{and} \quad \frac{N}{n^2} \mathbb{V}ar S_n^{(1)} \rightarrow 2 =: \tilde{\sigma}_1^2.$$

Similarly one can prove that

$$\frac{N}{n^2} \text{Cov}(S_n^{(0)}, S_n^{(1)}) \rightarrow -1.$$

Therefore for the correlation coefficients we have

$$\rho(S_n^{(0)}, S_n^{(1)}) \rightarrow -1. \quad (5.44)$$

Since

$$\frac{N}{n^2} \mathbb{V}\text{ar } S_n^{(2)} \rightarrow \frac{1}{2} = \sigma_2^{(2)} \quad \text{and} \quad \frac{N}{n^2} \text{Cov}(S_n^{(0)}, S_n^{(2)}) \rightarrow \frac{1}{2}$$

we also have

$$\rho(S_n^{(0)}, S_n^{(2)}) \rightarrow 1. \quad (5.45)$$

We consider the cases when  $r \geq 2$  and  $\lambda = \infty$  or  $r \geq 0$  and  $\lambda \in (0, \infty)$ .

**Theorem 8.** *Let  $N/n \rightarrow \lambda \in (0, \infty]$ . Let either*

(a)  $r \geq 0$  and  $\lambda < \infty$

or

(b)  $r \geq 2$ ,  $\lambda = \infty$  and  $\frac{n^r}{N^{r-1}} \rightarrow \infty$ .

Then

$$\frac{S_n^{(r)} - \mathbb{E} S_n^{(r)}}{\sqrt{n^r / N^{r-1}}} \xrightarrow{d} \mathcal{N}(0, \sigma_r^2).$$

*Proof.* Write  $c_{i,n}$  as

$$c_{i,n} = L_n^i e^{i \frac{n}{N}} \left(1 - \frac{i}{N}\right)^{n-ir} T(i, N) T(ir, n), \quad \text{where} \quad L_n = \frac{n^r e^{-\frac{n}{N}}}{r! N^{r-1}}.$$

Then, the representation (4.20) holds with

$$Q_{j+1}(i) = \left(r - \frac{j}{j+1} \frac{n}{N}\right) \left(\frac{n}{N}\right)^j i^{j+1} - \left(\frac{n}{N}\right)^j \mathcal{Q}_{j+1}(i) - \mathcal{Q}_{j+1}(ri).$$

Since  $\mathbb{E} S_n^{(r)}$ ,  $\mathbb{V}\text{ar } S_n^{(r)}$  and  $L_n$  are of order  $n^r / N^{r-1}$  the final result follows by Remark 6. □

Similarly as in the case of indistinguishable balls, using (5.45) and (5.44) we get

**Theorem 9.** *Let  $N/n \rightarrow \infty$  and  $n^2/N \rightarrow \infty$ . Then for  $r = 0, 1$*

$$\sqrt{N} \frac{S_n^{(r)} - \mathbb{E} S_n^{(r)}}{n} \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}_r^2).$$

### 5.2.3 Colored balls

An urn contains  $NM$  balls,  $M$  balls of each of  $N$  colors. From the urn a simple random sample of  $n$  elements is drawn. We want to study the asymptotics of the number of colors with exactly  $r$  balls in the sample. More precisely, let  $\xi_i = \xi_i^{(n)}$  denote the number of balls of color  $i$ ,  $i = 1, \dots, N$ . Then

$$\mathbb{P}(\xi_1 = k_1, \dots, \xi_N = k_N) = \frac{\prod_{i=1}^N \binom{M}{k_i}}{\binom{NM}{n}}$$

for all integers  $k_i \geq 0$ ,  $i = 1, \dots, N$ , such that  $\sum_{i=1}^N k_i = n$ . Obviously, the random vector  $(\xi_1, \dots, \xi_N)$  is exchangeable and the gas equation (5.26) holds with  $\eta_i \sim b(M, p)$ ,  $0 < p < 1$ .

For an integer  $r \geq 0$  we want to study the asymptotics of

$$S_n^{(r)} = \sum_{i=1}^N I(\xi_i = r).$$

For the  $i$ th factorial moments we get

$$c_{i,n} = (N)_i \mathbb{P}(\xi_1 = \dots = \xi_i = r) = (N)_i \frac{\binom{M}{r}^i \binom{(N-i)M}{n-ri}}{\binom{NM}{n}}. \quad (5.46)$$

Consequently, for any  $r = 0, 1, \dots$

$$\mathbb{E} S_n^{(r)} = c_{1,n} = N \frac{\binom{M}{r} \binom{(N-1)M}{n-r}}{\binom{NM}{n}}$$

and

$$\mathbb{V}ar S_n^{(r)} = c_{2,n} - c_{1,n}^2 + c_{1,n} = N(N-1) \frac{\binom{M}{r}^2 \binom{(N-2)M}{n-2r}}{\binom{NM}{n}} - N^2 \frac{\binom{M}{r}^2 \binom{(N-1)M}{n-r}^2}{\binom{NM}{n}^2} + N \frac{\binom{M}{r} \binom{(N-1)M}{n-r}}{\binom{NM}{n}}.$$

In the asymptotics below we consider the situation when  $n \rightarrow \infty$ ,  $\frac{N}{n} \rightarrow \lambda \in (0, \infty]$ , and  $M = M(n) \geq n$ .

Although the computational details are different, asymptotic formulas for  $\mathbb{E} S_n^{(r)}$ ,  $\mathbb{V}ar S_n^{(r)}$ ,  $\mathbb{C}ov(S_n^{(0)}, S_n^{(1)})$  and  $\mathbb{C}ov(S_n^{(0)}, S_n^{(2)})$  are literally the same as for their counterparts in the case of occupancy for distinguishable balls studied in Subsection 5.2.2.

First we will consider the limit result in the case  $r \geq 2$ ,  $\lambda = \infty$ , and  $r \geq 0$ ,  $\lambda \in (0, \infty)$ .

**Theorem 10.** *Let  $N/n \rightarrow \lambda \in (0, \infty]$  and  $M = M(n) \geq n$ . Let either*

(a)  $r \geq 0$  and  $\lambda < \infty$

or

(b)  $r \geq 2$ ,  $\lambda = \infty$  and  $\frac{n^r}{N^{r-1}} \rightarrow \infty$ .

Then

$$\frac{S_n^{(r)} - \mathbb{E} S_n^{(r)}}{\sqrt{n^r / N^{r-1}}} \xrightarrow{d} \mathcal{N}(0, \sigma_r^2).$$

*Proof.* Rewrite the formula (5.46) as

$$c_{i,n} = L_n^i \frac{T((M-r)i, NM-n) T(i, N) T(ri, n)}{T(Mi, NM)} \quad \text{with} \quad L_n = Nn^r \binom{M}{r} \frac{(1 - \frac{n}{NM})^M}{(NM-n)^r}.$$

Thus the representation (4.20) holds with

$$Q_{j+1}(i) = \left(\frac{n}{NM}\right)^j Q_{j+1}(Mi) - \left(\frac{n}{NM-n}\right)^j Q_{j+1}((M-r)i) - \left(\frac{n}{N}\right)^j Q_{j+1}(i) - Q_{j+1}(ri).$$

We need to see that the polynomials  $Q_j$  satisfy bound (4.21). This is clearly true for each of the last two terms in the above expression for  $Q_{j+1}$ . For the first two terms we have

$$\begin{aligned} & \left| \left(\frac{n}{NM}\right)^j Q_{j+1}(Mi) - \left(\frac{n}{NM-n}\right)^j Q_{j+1}((M-r)i) \right| \\ &= \left| \left(\frac{n}{NM}\right)^j \sum_{k=1}^{Mi-1} k^j - \left(\frac{n}{NM-n}\right)^j \sum_{k=1}^{(M-r)i-1} k^j \right| \\ &\leq \left| \left(\frac{n}{NM}\right)^j - \left(\frac{n}{NM-n}\right)^j \right| Q_{j+1}(Mi) + \left(\frac{n}{NM-n}\right)^j \sum_{k=(M-r)i}^{Mi-1} k^j. \end{aligned}$$

Since

$$\left| \left(\frac{n}{NM}\right)^j - \left(\frac{n}{NM-n}\right)^j \right| \leq \left(\frac{n}{NM-n}\right)^j \frac{jn}{NM} \leq \left(\frac{2n}{NM-n}\right)^j \frac{n}{NM},$$

and

$$\sum_{k=(M-r)i}^{Mi-1} k^j \leq rM^j i^{j+1},$$

and  $Q_{j+1}(Mi) \leq M^{j+1} i^{j+1}$  we conclude that  $Q_j$ 's do satisfy (4.21).

Clearly,  $\mathbb{E} S_n^{(r)}$ ,  $\mathbb{V}ar S_n^{(r)}$  and  $L_n$  are of order  $n^r/N^{r-1}$  and again we conclude the proof by referring to Remark 6.  $\square$

Asymptotic normality for  $S_n^{(1)}$  and  $S_n^{(0)}$  for  $\lambda = \infty$  also holds with the exact statement identical to Theorem 9 for distinguishable balls.

#### 5.2.4 Rooted trees in random forests

Let  $\mathcal{T}(N, n)$  denote a forest with  $N$  roots (that is  $N$  rooted trees) and  $n$  non-root vertices. Consider a uniform distribution on the set of such  $\mathcal{T}(N, n)$  forests. Let  $\xi_i = \xi_i^{(n)}$ ,  $i = 1, \dots, N$ , denote the number

of non-root vertices in the  $i$ th tree. Then (see for instance [3], [18] or [19]) for any  $k_i \geq 0$  such that  $\sum_{i=1}^N k_i = n$

$$\mathbb{P}(\xi_1 = k_1, \dots, \xi_N = k_N) = \frac{n!}{\prod_{i=1}^N k_i} \frac{\prod_{i=1}^N (k_i + 1)^{k_i - 1}}{N(N+n)^{n-1}}.$$

Note that this distribution is exchangeable and that it is a gas with  $\eta_i$  in (5.26) given by

$$\mathbb{P}(\eta_i = k) = \frac{\lambda^k (k+1)^{k-1}}{k!} e^{-(k+1)\lambda}, \quad k = 0, 1, \dots; \quad \lambda > 0.$$

We mention in passing that the distribution of  $\eta_i$  may be identified as Abel distribution discussed in [17] with (in their notation)  $p = 1$  and  $\theta = \ln \lambda - \lambda$ . We refer to Example D in [17] for more information on Abel distributions, including further references.

For a fixed number  $r \geq 0$  we are interested in the number  $S_n^{(r)}$  of trees with exactly  $r$  non-root vertices

$$S_n^{(r)} = \sum_{i=1}^N I(\xi_i = r).$$

Since the  $i$ th factorial moment of  $S_n^{(r)}$  is of the form

$$c_{i,n} = (N)_i \mathbb{P}(\xi_1 = \dots = \xi_i = r)$$

we have to find the marginal distributions of the random vector  $(\xi_1, \dots, \xi_N)$ . From the identity

$$s \sum_{k=0}^m \binom{m}{k} (k+1)^{k-1} (m-k+s)^{m-k-1} = (s+1)(m+1+s)^{m-1}$$

which is valid for any natural  $m$  and  $s$  we easily obtain that for  $k_j \geq 0$  such that  $\sum_{j=1}^{i+1} k_j = n$

$$\mathbb{P}(\xi_1 = k_1, \dots, \xi_i = k_i) = \frac{n!}{\prod_{j=1}^{i+1} k_j!} \frac{(N-i)(k_{i+1} + N-i)^{k_{i+1}-1} \prod_{j=1}^i (k_j + 1)^{k_j-1}}{N(N+n)^{n-1}}.$$

Therefore

$$c_{i,n} = \frac{(r+1)^{i(r-1)}}{(r!)^i} \frac{N-i}{N} \frac{(N)_i (n)_{ri}}{(n+N-(r+1)i)^{ri}} \left(1 - \frac{(r+1)i}{n+N}\right)^{n-1}. \quad (5.47)$$

Hence

$$\mathbb{E} S_n^{(r)} = c_{1,n} = \frac{(r+1)^{r-1}}{r!} \frac{(N-1)(n)_r}{(n+N-r-1)^r} \left(1 - \frac{r+1}{n+N}\right)^{n-1}.$$

Thus if  $N/n \rightarrow \lambda \in (0, \infty]$  we have

$$\frac{N^{r-1}}{n^r} \mathbb{E} S_n^{(r)} \rightarrow \frac{(r+1)^{r-1}}{r!} \left(\frac{\lambda}{\lambda+1}\right)^r e^{-\frac{r+1}{\lambda+1}} \left(= \frac{(r+1)^{r-1}}{r!} \text{ for } \lambda = \infty\right).$$

Since  $\mathbb{V}ar S_n^{(r)} = c_{2,n} - c_{1,n}^2 + c_{1,n}$  elementary but cumbersome computations lead to

$$\frac{N^{r-1}}{n^r} \mathbb{V}ar S_n^{(r)} \rightarrow \sigma_r^2$$

$$= \frac{e^{-\frac{r+1}{\lambda+1}}}{(r+1)!} \left( \frac{\lambda(r+1)}{\lambda+1} \right)^r \left[ 1 - \frac{e^{-\frac{r+1}{\lambda+1}}}{(r+1)!} \left( \frac{r+1}{\lambda+1} \right)^r \right] - \lambda \left( \frac{e^{-\frac{r+1}{\lambda+1}}(\lambda r - 1)}{(r+1)!(\lambda+1)} \right)^2 \left( \frac{\lambda(r+1)^2}{(\lambda+1)^2} \right)^r$$

for  $r \geq 2$  and  $\lambda \in (0, \infty]$  and for  $r = 0, 1$  and  $\lambda \in (0, \infty)$ . For  $r = 0, 1$  and  $\lambda = \infty$ , and  $n^2/N \rightarrow \infty$  we have

$$\frac{N}{n^2} \mathbb{V}ar S_n^{(0)} \rightarrow \frac{3}{2} = \tilde{\sigma}_0^2 \quad \text{and} \quad \frac{N}{n^2} \mathbb{V}ar S_n^{(1)} \rightarrow 6 = \tilde{\sigma}_1^2.$$

Similarly one can prove that

$$\frac{N}{n^2} \mathbb{C}ov(S_n^{(0)}, S_n^{(1)}) \rightarrow -3.$$

Therefore for the correlation coefficients we have

$$\rho(S_n^{(0)}, S_n^{(1)}) \rightarrow -1. \quad (5.48)$$

Since

$$\frac{N}{n^2} \mathbb{V}ar S_n^{(2)} \rightarrow \frac{3}{2} = \sigma_2^{(2)} \quad \text{and} \quad \frac{N}{n^2} \mathbb{C}ov(S_n^{(0)}, S_n^{(2)}) \rightarrow 1$$

we also have

$$\rho(S_n^{(0)}, S_n^{(2)}) \rightarrow 1. \quad (5.49)$$

**Theorem 11.** Let  $N/n \rightarrow \lambda \in (0, \infty]$ . Let either

(a)  $r \geq 0$  and  $\lambda < \infty$

or

(b)  $r \geq 2$ ,  $\lambda = \infty$  and  $\frac{n^r}{N^{r-1}} \rightarrow \infty$ .

Then

$$\frac{S_n^{(r)} - \mathbb{E} S_n^{(r)}}{\sqrt{n^r/N^{r-1}}} \xrightarrow{d} \mathcal{N}(0, \sigma_r^2).$$

*Proof.* Since the asymptotics of  $\mathbb{V}ar S_n^{(r)}$  and  $\mathbb{E} S_n^{(r)}$  is of the same order as in Theorem 5.2.2 the conditions (2.7) and (2.8) are satisfied. Using (5.47) we write

$$c_{i,n} = L_n^i e^{i \frac{(n-1)(r+1)}{n+N}} \left( 1 - \frac{(r+1)i}{n+N} \right)^{n-1-ri} T(i+1, N) T(ri, n),$$

where

$$L_n = \frac{N(r+1)^{r-1}}{r!} \left( \frac{n}{n+N} \right)^r e^{-\frac{(n-1)(r+1)}{n+N}}.$$

Thus the representation (4.20) holds true with

$$Q_{j+1}(i) = \left( r - \frac{j(r+1)(n-1)}{(j+1)(n+N)} \right) \left( \frac{(r+1)(n-1)}{n+N} \right)^j i^{j+1} - \left( \frac{n}{N} \right)^j \mathcal{Q}_{j+1}(i+1) - \mathcal{Q}_{j+1}(ri).$$

The final result follows again by Remark 6 on noting that  $\mathbb{E} S_n^{(r)}$ ,  $\mathbb{V}ar S_n^{(r)}$  and  $L_n$  are of order  $n^r/N^{r-1}$ .  $\square$

Again, as in previous cases we use (5.48) and (5.49) to arrive at

**Theorem 12.** *Let  $N/n \rightarrow \infty$  and  $n^2/N \rightarrow \infty$ . Then for  $r = 0, 1$*

$$\sqrt{N} \frac{S_n^{(r)} - \mathbb{E} S_n^{(r)}}{n} \xrightarrow{d} \mathcal{N}(0, \tilde{\sigma}_r^2) .$$

### 5.3 Asymptotics in generalized inverse allocation schemes (gias)

Our final two settings are examples of the gas as defined in (5.27). Similarly to the case of gas for  $S_n^{(r)} = \sum_{i=1}^N I(\xi_i^{(n)} = r)$  we will obtain asymptotic normality of

$$\frac{S_n^{(r)} - \mathbb{E} S_n^{(r)}}{\sqrt{n}},$$

when  $N/n \rightarrow \lambda \in (0, \infty)$ .

#### 5.3.1 Exchangeable negative multinomial model

Let  $(\xi_i) = (\xi_i^{(n)})$  be a random vector with a negative multinomial distribution, that is

$$\mathbb{P}(\xi_1 = k_1, \dots, \xi_N = k_N) = \frac{(n + \sum_{j=1}^N k_j)!}{n! \prod_{j=1}^N k_j!} p^{\sum_{j=1}^N k_j} (1 - Np)^n.$$

Note that this is an exchangeable case of a model usually referred to as Bates–Neyman model introduced in [1]. We refer to [9, Chapter 36, Sec. 1–4] for a detailed account of this distribution, its properties, applications, and further references. Here, we just note that this is a gas for which (5.27) holds with  $\eta_0 \sim \text{Poisson}(\lambda_0)$ ,  $\eta_i \sim \text{Poisson}(\lambda_1)$ ,  $i = 1, \dots, N$ , and  $p = \frac{\lambda_1}{\lambda_0 + N\lambda_1}$ .

For a fixed integer  $r$  we are interested in the asymptotics of

$$S_n^{(r)} = \sum_{j=1}^N I(\xi_j = r).$$

Denoting  $\beta_n = (Np)^{-1} - 1$  we obtain

$$c_{i,n} = (N)_i \frac{(n + ri)!}{n!(r!)^i} \frac{(N\beta_n)^n}{(N\beta_n + i)^{n+ri}}. \quad (5.50)$$

To study the asymptotic properties of  $S_n^{(r)}$  we will assume that  $N/n \rightarrow \lambda \in (0, \infty)$ . Moreover we let  $p = p_n$  depend on  $n$  in such a way that  $Np_n \rightarrow \alpha \in (0, 1)$ , i.e.  $\beta_n \rightarrow \alpha^{-1} - 1$ . Consequently, setting  $\Delta := \frac{\alpha}{\lambda(1-\alpha)}$ , for any  $r = 0, 1, \dots$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} S_n^{(r)} = \frac{\lambda \Delta^r}{r!} e^{-\Delta}$$



and

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{V}ar S_n^{(r)} = \frac{\lambda \Delta^r e^{-\Delta}}{r!} \left[ 1 - \frac{\Delta^r e^{-\Delta}}{r!} (1 - \lambda(r - \Delta)^2) \right] =: \sigma_r^2. \quad (5.51)$$

**Theorem 13.** Let  $N/n \rightarrow \lambda \in (0, \infty)$  and  $Np_n \rightarrow \alpha \in (0, 1)$ . Then for any  $r = 0, 1, \dots$

$$\frac{S_n^{(r)} - \mathbb{E} S_n^{(r)}}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_r^2),$$

with  $\sigma_r^2$  defined in (5.51).

*Proof.* Write (5.50) as

$$c_{i,n} = L_n^i e^{i \frac{n}{N\beta_n}} \frac{T(ri + 1, -n) T(i + 1, N)}{\left(1 + \frac{i}{N\beta_n}\right)^{n+ri}} \quad \text{with} \quad L_n = \frac{n^r e^{-\frac{n}{N\beta_n}}}{r! N^{r-1} \beta_n^r}.$$

Thus the representation (4.20) holds true with

$$Q_{j+1}(i) = \left(r - \frac{jn}{(j+1)N\beta_n}\right) \left(-\frac{n}{N\beta_n}\right)^j i^{j+1} + (-1)^{j+1} Q_{j+1}(ri + 1) - \left(\frac{n}{N}\right)^j Q_{j+1}(i).$$

Moreover,  $\mathbb{E} S_n^{(r)}$ ,  $\mathbb{V}ar S_n^{(r)}$  and  $L_n$  are all of order  $n$  and thus the final conclusion follows from Remark 4.  $\square$

### 5.3.2 Dirichlet negative multinomial model

Finally we consider exchangeable version of what is known as Dirichlet model of buying behavior introduced in a seminal paper by Goodhardt, Ehrenberg and Chatfield [4] and subsequently studied in numerous papers up to the present days. This distribution is also mentioned in [9] (see Chapter 36, Sec. 6 of that book). Writing, as usually,  $\xi_i^{(n)} = \xi_i$ , the distribution under study has the form

$$\mathbb{P}(\xi_1 = k_1, \dots, \xi_N = k_N) = \frac{\left(n + \sum_{i=1}^N k_i\right)!}{n! \prod_{j=1}^N k_j!} \frac{\Gamma(Na + b)}{\Gamma^N(a)\Gamma(b)} \frac{\Gamma(n + b) \prod_{j=1}^N \Gamma(k_j + a)}{\Gamma\left(Na + b + n + \sum_{j=1}^N k_j\right)}$$

for any  $k_i = 0, 1, \dots, i = 1, \dots, N$ . Here  $n > 0$  is an integer and  $a, b > 0$  are parameters. This is again a gias with  $\eta_i \sim nb(a, p)$ ,  $i = 1, \dots, N$  and  $\eta_0 \sim nb(b, p)$ ,  $0 < p < 1$ . When  $a$  and  $b$  are integers we recall a nice interpretation of  $(\xi_1, \dots, \xi_N)$  via the Polya urn scheme: An urn contains  $b$  black balls and  $a$  balls in each of  $N$  non-black colors. In subsequent steps a ball is drawn at random and returned to the urn together with one ball of the same color. The experiment is continued until the  $n$ th black ball is drawn. Then  $\xi_i$  is the number of balls of the  $i$ th color at the end of experiment,  $i = 1, \dots, N$ . This distribution can also be viewed as multivariate negative hypergeometric law of the second kind.

From the fact that  $c_{i,n} = (N)_i \mathbb{P}(\xi_1 = \dots = \xi_i = r)$  we get

$$c_{i,n} = (N)_i \frac{(n + ri)!}{n! (r!)^i} \frac{\Gamma(ia + b)}{\Gamma^i(a)\Gamma(b)} \frac{\Gamma(n + b) \Gamma^i(r + a)}{\Gamma((r + a)i + n + b)}. \quad (5.52)$$

To study the asymptotic behavior of  $S_n^{(r)}$  we will assume that  $N/n \rightarrow \lambda \in (0, \infty)$  and that  $b = b_n$  depends on  $n$  in such a way that  $b_n/n \rightarrow \beta > 0$ .

Below we use the following product representation of Gamma function

$$\Gamma(x) = \frac{1}{xe^{\gamma x}} \prod_{k \geq 1} \frac{e^{\frac{x}{k}}}{1 + \frac{x}{k}}, \quad (5.53)$$

where  $\gamma$  is the Euler constant and  $x > 0$ . We also recall (see e.g. [23, Section 12.16]) that for a digamma function  $\Psi(x) = \frac{d}{dx} \ln(\Gamma(x))$  we have

$$\Psi(x+1) = -\gamma + \sum_{k \geq 1} \left( \frac{1}{k} - \frac{1}{k+x} \right) \quad x \neq -1, -2, \dots$$

Then for any  $0 < x < y$  we can write

$$\frac{\Gamma(x+y)}{\Gamma(y)} = e^{x\Psi(y+1)} \frac{y}{x+y} e^{\sum_{k \geq 1} \left( \frac{x}{k+y} - \log\left(1 + \frac{x}{k+y}\right) \right)} \quad (5.54)$$

and the series  $\sum_{k \geq 1} \left( \frac{x}{k+y} - \log\left(1 + \frac{x}{k+y}\right) \right)$  converges. Note that,

$$\Psi(y) - \ln y \rightarrow 0 \quad \text{as } y \rightarrow \infty \quad (5.55)$$

so that, if  $\alpha_n/n \rightarrow \alpha$  then for any  $x > 0$

$$n^{-x} \frac{\Gamma(\alpha_n + x)}{\Gamma(\alpha_n)} \rightarrow \alpha^x.$$

Consequently,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} S_n^{(r)} = \frac{\lambda \Gamma(a+r)}{r! \Gamma(a)} \frac{\beta^a}{(1+\beta)^{a+r}}.$$

Similarly,

$$\lim_{n \rightarrow \infty} \text{Var } S_n^{(r)} = \frac{\lambda \Gamma(r+a) \beta^a}{r! \Gamma(a) (1+\beta)^{a+r}} \left( 1 - \frac{\Gamma(r+a) \beta^a \left[ 1 + \lambda \left( \frac{(a+r)^2}{\beta+1} - \frac{a^2}{\beta} - r^2 \right) \right]}{r! \Gamma(a) (1+\beta)^{a+r}} \right) =: \sigma_r^2. \quad (5.56)$$

**Theorem 14.** Let  $N/n \rightarrow \lambda \in (0, \infty)$  and  $b_n/n \rightarrow \beta \in (0, \infty)$ . Then for any  $r = 0, 1, \dots$

$$\frac{S_n^{(r)} - \mathbb{E} S_n^{(r)}}{\sqrt{n}} \xrightarrow{d} \mathcal{N}(0, \sigma_r^2),$$

with  $\sigma_r^2$  defined in (5.56).

*Proof.* Note that (5.52) can be written as

$$c_{i,n} = \left( \frac{N \Gamma(r+a) n^r}{r! \Gamma(a)} \right)^i T(ir+1, -n) T(i, N) \frac{\Gamma(b_n + ia)}{\Gamma(b_n)} \frac{\Gamma(n + b_n)}{\Gamma(n + b_n + i(r+a))}.$$

Moreover, setting

$$h_j(x) = \sum_{k \geq 1} \frac{1}{(k+x)^{j+1}}, \quad x > 0, \quad j \geq 2.$$

we see that (5.54) can be developed into

$$\frac{\Gamma(x+y)}{\Gamma(y)} = e^{x\Psi(y+1)} e^{\sum_{j \geq 1} \left( \frac{(-x)^j}{j y^j} + \frac{(-x)^{j+1}}{j+1} h_j(y) \right)}.$$

Therefore, taking  $(x, y) = (ia, b_n)$  and  $(x, y) = (i(r+a), n+b_n)$  we decompose  $c_{i,n}$  according to (4.20) where

$$L_n = \frac{N\Gamma(r+a)n^r}{r!\Gamma(a)} e^{a\Psi(1+b_n) - (r+a)\Psi(1+n+b_n)}$$

and

$$\begin{aligned} Q_{j+1}(i) &= \left[ \left( \frac{n(r+a)}{b_n+n} \right)^j - \left( \frac{na}{b_n} \right)^j \right] (-1)^{j+1} i^j + \frac{j(-n)^j}{j+1} \left[ (r+a)^{j+1} h_j(b_n+n) - a^{j+1} h_j(b_n) \right] i^{j+1} \\ &\quad + (-1)^{j+1} Q_{j+1}(ir+1) - \left( \frac{n}{N} \right)^j Q_{j+1}(i). \end{aligned}$$

On noting that  $\alpha_n/n \rightarrow \alpha$  implies that  $n^j g_j(\alpha_n) < c^j(\alpha)$  uniformly in  $n$  we conclude that polynomials  $Q_j$  satisfy condition (4.21). Moreover, (5.55) yields that  $L_n$  is of order  $n$ . Since  $\mathbb{E} S_n^{(r)}$  and  $\mathbb{V}ar S_n^{(r)}$  are of order  $n$  too, the result follows by Remark 4.  $\square$

## References

- [1] BATES, G. E., NEYMAN, J. Contributions to the theory of accident proneness I. An optimistic model of the correlation between light and severe accidents. *Univ. California Publ. Statist.* **1** (1952), 215–253.
- [2] BILLINGSLEY, P. *Probability and Measure*. Wiley (3rd edition), New York 1995.
- [3] CHUPRUNOV, A., FAZEKAS, I. An inequality for moments and its applications to the generalized allocation scheme. *Publ. Math. Debrecen*, **76** (2010), 271–286.
- [4] GOODHART, G.J., EHRENBERG, A.S.C., CHATFIELD, C. The Dirichlet: A comprehensive model of buying behavior. *Journal of the Royal Statistical Society, Section A*, **147** (part 5) (1984), 621–655.
- [5] GRAHAM, R.L., KNUTH, D.E., PATASHNIK, O. *Concrete Mathematics*. Addison-Wesley Publishing Company, Reading 1994.
- [6] HALDANE, J. B. S. On a method of estimating frequencies. *Biometrika* **33** (1945), 222–225.
- [7] HWANG, H-K., JANSON, S. Local limit theorems for finite and infinite urn models. *Ann. Probab.*, **36** (2008), 992–1022.

- [8] JANSON, S. ŁUCZAK, T. RUCINSKI, A., *Random Graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York 2000.
- [9] JOHNSON, N. L., KOTZ, S., BALAKRISHNAN, N. *Multivariate Discrete Distributions*. Wiley Series in Probability and Statistics, Wiley-Interscience, New York 1997.
- [10] JOHNSON, N. L., KOTZ, S., KEMP, A. W. *Univariate Discrete Distributions*, 2nd ed. Wiley Series in Probability and Statistics, Wiley-Interscience, New York 1992.
- [11] KEENER, R. W., WU, W. B., On Dirichlet multinomial distributions. in: *Random walk, sequential analysis and related topics*, pp. 118–130. World Sci. Publ., Hackensack, NJ, 2006.
- [12] KENDALL, M.G., STUART, A. *The Advanced Theory of Statistics. Vol. 1: Distribution Theory*. Addison-Wesley Publ. Comp., Reading 1969.
- [13] KOLCHIN, V. F. A certain class of limit theorems for conditional distributions. *Litovsk. Mat. Sb.* **8** (1968), 53–63.
- [14] KOLCHIN, V. F. *Random Graphs*. Encyclopedia of Mathematics and its Applications, vol. 53. Cambridge Univ. Press, Cambridge 1999.
- [15] KOLCHIN, V. F. *Random Mappings*. Optimization Software Inc. Publications Division, New York 1986.
- [16] KOLCHIN, V.F., SEVASTYANOV, B.A., CHISTYAKOV, V.P. *Random Allocations*. Winston & Sons, Washington 1978.
- [17] LETAC, G., MORA, M. Natural real exponential families with cubic variance functions. *Ann. Statist.* **18** (1990), 1 – 37.
- [18] PAVLOV, YU. L. Limit theorems for the number of trees of a given size in a random forest. *Mat. Sbornik* **103 (145)** (1977), 335 – 345.
- [19] PAVLOV, YU. L. *Random forests*. VSP Utrecht 2000.
- [20] STANLEY, R.P. *Enumerative Combinatorics, Vol. 2*. Cambridge Univ. Press, Cambridge 1999.
- [21] STANLEY, R.P. Personal communication, (2009).
- [22] VAN DER VAART, A.W. *Asymptotic Statistics*. Cambridge University Press, Cambridge 1998.
- [23] WHITTAKER, E. T., WATSON, G. N., *A Course of Modern Analysis* (4th ed. reprinted). Cambridge Univ. Press. Cambridge 1996.